Genome-wide view of genetic diversity reveals paths of selection and cultivar differentiation in peach domestication.

AUTHOR(S):

Akagi, Takashi; Hanada, Toshio; Yaegaki, Hideaki; Gradziel, Thomas M.; Tao, Ryutaro

CITATION:

Akagi, Takashi ...[et al]. Genome-wide view of genetic diversity reveals paths of selection and cultivar differentiation in peach domestication.. DNA Research 2016, 23(3): 271-282

ISSUE DATE:

2016-04-15

URL:

http://hdl.handle.net/2433/216923

DNA Research, 2016, 23(3), 271–282
doi: 10.1093/dnares/dsw014
Advance Access Publication Date: 15 April 2016
Full Paper

Full Paper

Genome-wide view of genetic diversity reveals paths of selection and cultivar differentiation in peach domestication

Takashi Akagi[1,*], Toshio Hanada[2], Hideaki Yaegaki[3], Thomas M. Gradziel[4], and Ryutaro Tao[1]

[1]Laboratory of Pomology, Graduate School of Agriculture, Kyoto University, Kita-shirakawa, Oiwake-cho, Kyoto 606-8502, Japan, [2]Apple Research Division, NARO Institute of Fruit Tree Science, Morioka 020-0123, Japan, [3]Breeding and Pest Management Division, NARO Institute, Tsukuba, Ibaragi 305-8605, Japan, and [4]Department of Plant Sciences, University of California Davis, CA 95616, USA

*To whom correspondence should be addressed. Tel. +81 75-753-6051. E-mail: takashia@kais.kyoto-u.ac.jp

Edited by Dr Masahiro Yano

Received 20 November 2015; Accepted 2 March 2016

Abstract

Domestication and cultivar differentiation are requisite processes for establishing cultivated crops. These processes inherently involve substantial changes in population structure, including those from artificial selection of key genes. In this study, accessions of peach (*Prunus persica*) and its wild relatives were analysed genome-wide to identify changes in genetic structures and gene selections associated with their differentiation. Analysis of genome-wide informative single-nucleotide polymorphism loci revealed distinct changes in genetic structures and delineations among domesticated peach and its wild relatives and among peach landraces and modern fruit (F) and modern ornamental (O-A) cultivars. Indications of distinct changes in linkage disequilibrium extension/decay and of strong population bottlenecks or inbreeding were identified. Site frequency spectrum- and extended haplotype homozygosity-based evaluation of genome-wide genetic diversities supported selective sweeps distinguishing the domesticated peach from its wild relatives and each F/O-A cluster from the landrace clusters. The regions with strong selective sweeps harboured promising candidates for genes subjected to selection. Further sequence-based evaluation further defined the candidates and revealed their characteristics. All results suggest opportunities for identifying critical genes associated with each differentiation by analysing genome-wide genetic diversity in currently established populations. This approach obviates the special development of genetic populations, which is particularly difficult for long-lived tree crops.

Key words: artificial selection, cultivar differentiation, domestication, linkage disequilibrium, tree crop

1. Introduction

Plant domestication establishes a vital co-dependence between humans and plants.[1] Previous studies of the plant domestication process have focused mainly on seed-propagated annual (herbaceous) species, often with the goal of identifying wild progenitors, as well as changes in genetic structures associated with domestication.[2–5] The results have provided historical insights into the genetic adaptation required for domestication, which has relevance to crop breeding strategies. During the domestication and early breeding process, crops typically experience population bottlenecks, including extensive artificial selection for improved crop quality and local adaptation.[6] Evidence of this selection remains in the patterns of genetic diversity within cultivated

genomes,[7] as characterized for major crops such as rice,[8,9] wheat,[10] and maize.[11–14] These studies also identified important genomic regions and/or genes associated with artificial selection. Genes under strong selection reflect the main driving forces for each historical differentiation. Thus, their detection can provide valuable targets for further crop improvement research,[15] as well as opportunities to improve current genomic selection (GS) models.[16] Conventional mapping approaches such as genotype–phenotype association mapping, though effective for identifying loci contributing to specific traits, are not as applicable to diversified populations with their large variety of genetic structures.[15] For approaches using linkage mapping, perennial plants present additional challenges because of their large size and long generation time. Thus, genome-wide association study (GWAS)[17,18] used in conjunction with analysis of the effect of selection on the genomes of current populations may be a good strategy for identifying critical loci defining perennial crops.

In contrast to those for annual crops, studies depicting paths of domestication for perennial plants, including tree crops, are rare.[18–22] Trees have unique features affecting their manner and rate of domestication: (i) individual genotypes can be maintained by vegetative or clonal propagation, (ii) trees have a long generation time, frequently with a long juvenile phase, and (iii) wild populations are often large with high levels of gene flow.[23–25] Additionally, different reproductive systems and cultural practices among species may have led to substantially different effects on domestication and breeding pathways. A typical example for reproductive systems is self-incompatibility (SI), which promotes outcrossing. Outcrossing generally disturbs the fixation of selected alleles and the resulting domestication.[26] Rowlands hypothesized that important crops undergo breakdown of SI during domestication and/or breeding, owing to artificial selection for stable production in more homozygous lines.[27] It has been proposed[22] that the domesticated apple did not experience a population bottleneck in the domestication step, presumably because of its long generation time, the nature of the SI system, and breeding practices such as selection from open-pollinated seed (leading to 'chance seedlings'). Similarly, domesticated grapevine shows only a weak bottleneck, partially because of specific cultural practices including the maintenance of many old varieties with wide genetic diversity.[18] It has been proposed that tree crops are, in general, less affected by population bottlenecks and other selection limitations during the domestication process.[21,22]

Peach [*Prunus persica* (L.) Batsch] is one of the most important tree crops worldwide. It is in the Rosaceae family, which also includes apricot, almond, plum, strawberry, rose, pear, and apple, and has been reported to have been cultivated for >4,000 yrs.[28] Archaeological findings support the domestication of peach and its artificial selection for various desirable traits in China from as early as 1000 BC,[29] although there are only a few reports on specific origins or required traits for domestication.[30] Modern varieties of cultivated peach are conventionally categorized into two main groups: fruit varieties and flower-ornamental varieties, though this grouping does not consider genetic structure. For fruit varieties, many regional cultivars exist. However, most modern cultivars in Europe and North America appear to have originated from only a few old cultivars used in early twentieth-century North American breeding programs.[28,31] Based on molecular marker data, cv. Chinese Cling has been proposed as an important founding parent in these early breeding programs.[32,33] In Asia, cv. Chinese Cling has also been proposed to be a founding parent of numerous modern cultivars,[34–36] though it should be mentioned that this designation may refer to multiple independent landraces and local selections.[37] The exact identities of many old cultivars and classification names reported in the older literature remain confusing given that they have been cultivated in China for thousands of years, often with multiple introductions to Western countries.[38] For the flower-ornamental peach group, little is known about the genetic structure and genetic diversity, except for some limited cultivars and landraces.[30]

Artificial selection is expected to occur not only during domestication but also in subsequent breeding for specific objectives. In domesticated chicken, genome-wide comparison of genetic diversity during recent breeding differentiation to layers or broilers revealed multiple independent targets of artificial selection.[39] The recent availability of genome-wide scans of plant species' DNA polymorphisms allows for a more detailed study of their population genetics, including characterization of linkage disequilibrium (LD) as well as genetic diversity, including pi-values for the detection of targets of selection during plant domestication.[9,10,30,40–42] The objective of this study was to better characterize the genetic structure of peach and to identify important selection events during domestication and subsequent cultivar differentiation, using a genome-wide single-nucleotide polymorphism (SNP) chip-based approach.

2. Materials and methods

2.1. Sample collection and genome-wide genotyping

For genome-wide genotyping, leaves were collected from 67 accessions of domesticated peach (*P. persica*; subg. *Amygdalus*), 20 accessions of peach relatives (four *Prunus davidiana*, one *P. kansuensis*, 12 *P. mira*, one *P. tangutica*, and two *P. webbii*; subg. *Amygdalus*), and 8 accessions of outgroup species [two Japanese plum (*P. salicina*; subg. *Prunus*), two Japanese apricot (*P. mume*; subgen. *Prunus*), and four sweet cherry (*P. avium*; subg. *Cerasus*)], from the UC Davis and USDA *Prunus* germplasm collections in Winters, California, USA, the NIFTS *Prunus* germplasm collections, Tsukuba, Japan, the Research Institute for Agriculture Okayama Prefectural Technology Center for Agriculture, Forestry, and Fisheries, Akaiwa, Japan, and the experimental orchard of Kyoto University, Kyoto, Japan, as summarized in Supplementary Table S1. Tissues from a total of 95 accessions were subjected to DNA extraction using Nucleon PhytoPure (GE Healthcare, Tokyo, Japan) and thereafter phenol/chloroform extraction.

Genotype calling was performed with an Illumina infinium peach 9K SNP chip, which was defined from a total of 1,022,354 SNPs that were identified from the resequencing data in a wide variety of 56 accessions in *Amygdalus*, mainly including peach (*P. persica*) and almond (*Prunus dulcis*), and ~75% of genic SNPs were verified in the peach genome.[43,44] We assayed 7,873 SNPs in this study. Each 5,180 and 6,605 informative SNP was selected for structural analysis and evaluation of LD and selective sweeps, respectively (see below).

2.2. Allele pruning

SNPs showing >5% missing data or <0.05 minor allele frequency (MAF) in domesticated peach cultivars were pruned with PLINK.[45] After filtering, 6,605 SNPs remained for use in estimating LD. SNP pairs showing strong LD were further pruned by defining a window of 50 SNPs, removing one of a pair of SNPs if $R^2 > 0.5$ (VIF threshold values = 2), and then shifting the window by three SNPs and repeating the procedure using PLINK. After filtering, 5,180 SNPs remained for use in the analyses of population structure.

2.3. Analysis of population structure

For the topology of the evolutionary tree, we aligned 5,180 concatenated SNPs to give two aligned positions for each SNP locus to reflect the diploid allelic states. The aligned concatenated sequences were

subjected to neighbour joining (NJ) using version 5.05 under the Poisson matrix with gamma distribution for the rates and with 1,000 bootstrap replications. Principal component analysis (PCA) was performed using the same 5,180 SNP set, using prcomp implemented in R version 2.15.3. Heterozygosity [expected (H_E) and observed (H_O)], inbreeding coefficient (F_{IS}), and Weir & Cockerham *F* statistics (F_{ST}) were calculated with GENEPOP 4.0.[46,47] Identity by descent (IBD) proportions were calculated as pi-hat values with PLINK, considering all pairings of the 67 accessions of domesticated peach (*P. persica*). We used the pruned 5,180 SNPs for the calculation. We had no calibration to infer first- or second-degree relationships from the information of actual pedigrees owing to the lack of information on reliable pedigrees and to the small number of accessions.

To evaluate delimitations in population structure, we performed individual-based Bayesian clustering with Markov chain Monte Carlo (MCMC) simulations, using STRUCTURE 2.2[48] and InStruct[49] to infer the population ancestry of genotypes in *K* predefined clusters. *K* values ranging from 2 to 10 were evaluated for subdivision of the full (domesticated or wild) peach population ($n = 87$). We fixed *K* at 2 for characterizing proportions of ancestry from two predefined ancestral gene pools in some combinations of the varietal complexes or of the resulting clusters, to support population subdivision, and to infer genetic introgression among the complexes according to Cornille et al.[22] We performed an independent test of each *K* using at least 200,000 MCMC iterations after 50,000 burn-in iterations. To evaluate inference of *K*, the model with the highest ln Pr(*K*)[48] and the Δ*K* model with the greatest second-order rate of change in ln Pr(*K*)[50] were examined.

Pairwise LD among the SNPs were calculated with PLINK, using 6,605 SNP sets considering MAF (<0.05) and missing reads (>20%) in the full population as candidates. We exploited all R^2 values in all pairs of SNPs in 10,000-kb windows. The R^2 values were independently calculated for each subpopulation for each chromosome. The LD decay distances in each domesticated peach subpopulation, in the whole domesticated peach population, and in a wild related population were defined as the first points at which the average R^2 values in a 100-kb bin revealed no significant difference ($P > 0.1$) against the background R^2 values, which were calculated from average intra-chromosomal comparisons.

2.4. Haplotype phasing

We characterized haplotypes of SNPs in each chromosome with fastPHASE v.1.2 using 6,605 SNP sets considering MAF and missing rates in the entire population.[49] In advance and against the full population, including 87 wild or domesticated accessions, we estimated imputation error rates using 1,583 SNPs on chromosome 4 and the following options in fastPHASE: number of random starts of the EM algorithm = 10, EM iterations = 25, lower limit numbers of clusters = 1, upper limit numbers of clusters = 20, and interval between values for number of clusters = 1. This analysis gave an imputation error rate of 0.0822 with $K = 8$ as an optimal condition, only slightly larger than the values reported for phasing of SNPs on grapevine chromosome 8 and in humans.[18,51] We adopted $K = 8$ and constructed haplotypes for each chromosome. Construction of haplotypes was also performed using each subpopulation defined in structure analysis (see Results section), which often showed a lower imputation error rate (*ca.* 0.06–0.17) than the full population.

2.5. Detection of selective sweeps

The 6,605 SNPs considering MAF and missing rates were used for three approaches, site frequency spectrum (SFS)-, integrated Haplotype Score (iHS)-, and XP-extended haplotype homozygosity (EHH)-based methods. The SFS-based approach was applied using the pooled heterozygosity (H_p)[39] of each cluster in a 400-kb sliding window with a 100-kb step. Windows contained only four or fewer positions where SNPs were removed. All SNP alleles were completely given in bi-allelic states, and thus, the pooled heterozygosity was given by the average expected heterozygosity in each window, as follows:

$$H_p = \frac{1}{S}\sum_{m=1}^{S}\left(\frac{2N(m)_{\mathrm{maj}}N(m)_{\mathrm{min}}}{\{N(m)_{\mathrm{maj}} + N(m)_{\mathrm{min}}\}^2}\right)$$

Here, *S* is the number of SNP positions in a window, $N(m)_{\mathrm{maj}}$ and $N(m)_{\mathrm{min}}$ are the numbers of major and minor alleles, respectively, in *m*th SNP locus in a window. Individual H_p values were then *Z*-transformed as follows: $ZH_p = (H_p - \mu H_p)/\sigma H_p$, according to a report on pooled heterozygosity for selective sweep analysis in chicken.[39] The allele information was used in an unphased state.

Both iHS and XP-EHH tests were based on EHH, which is defined as the probability that two randomly chosen chromosomes carrying the core haplotype of interest are identical by descent (as assayed by homozygosity at all SNPs) for the entire interval from the core region to point *x*.[52] We detected EHH in each subpopulation using the program Sweep.[53] We approached the computation of the integral of observed EHH (iHH)[54] in 1,000 kb from each SNP core by measuring EHH in every 100-kb bin. EHH decayed to below 0.05 by 1,000 kb from SNP cores in most cases.

For detection of iHS, iHH values from two core alleles at one SNP core position in one population were defined as iHH_A and iHH_D, which originally corresponded to ancestral and derived alleles, respectively.[54] iHS statistic is then given as unstandardized $iHS = \ln(iHH_A/iHH_D)$ (1).[54] Finally, we obtained the standardized iHS as a *Z*-transformed value of the unstandardized iHS. We examined the reciprocal states of two core alleles (iHH_A and iHH_D) because of the sparseness of information on the derivation of the core alleles among peach subpopulations. XP-EHH is also obtained by Equation (1), focusing on the same core alleles in comparison to two populations.[53] For two populations, A and B, the log values of the integral EHH, I_A and I_B (like iHH_A and iHH_D in the iHS test), $\ln(I_A/I_B)$ give an index of selection specific to either of the two populations. An unusually positive value of $\ln(I_A/I_B)$ suggests selection in population A, whereas a negative value suggests selection in population B. The standardized XP-EHH is given as a *Z*-transformed value of $\ln(I_A/I_B)$, the details of which are presented by Sabeti.[53] The standardized iHS and XP-EHH were transformed to *P* values using R for graphical plotting.

2.6. Sequencing and genetic diversity analysis in selected regions

To exploit genes under selection during cultivar differentiation, full lengths of a total of 15 genes located on the region under putative selection were amplified by PCR using 30 peach accessions (Supplementary Table S3a) and then sequenced with an Illumina HiSeq 2000 (Illumina) as paired-end 100 (PE100). The libraries were constructed based on in-house-developed protocols described previously.[55] Approximately 4% of a sequencing lane was dedicated to all samples to yield at least 100× coverage in any region. Over 15% of read coverage for each SNP was considered as informative. All bioinformatic and statistical analyses were performed on local servers at the UC Davis Genome Center (Davis, CA, USA). Raw reads without adapter sequences were subjected to trimming (length > 35 bp, mean sliding window of 5 bp phred quality score ≥20) using custom Python scripts. Reads were then mapped to the reference sequences from the peach

genome (Peach v1.0) using the Burrows–Wheeler Aligner (BWA) tools.[56] Informative SNPs were identified using Sequence Alignment/Map (SAM) tool[57] and custom Python scripts. Informative SNP strings and statistics for selective pressure (nucleotide diversity (π) and Tajima's *D*) were evaluated with DnaSP 5.1 using sequence alignments constructed with ClustalX version 2.0 with minor revision using SeaView version 4, according to previous studies.[8,12,58,59]

3. Results

3.1. Genotyping and defining the genetic structure

The topology of the evolutionary tree, constructed using 5,180 concatenated SNPs, showed distinct differentiation between peach and its wild relatives and also among the cultivar complexes, based on classical classifications by morphology or use (Fig. 1). Six major clusters were defined in the evolutionary tree: wild species (W), landraces (L), modern fruit cultivars (F), and modern ornamental cultivars (O-A and O-B for only ornamental usage and FO for fruit and ornamentals). The F (including FO) and O-A clearly diverged from landraces with strong statistical support (with bootstrap values of 70/100 and 74/100, respectively). The L clade was divided into two clades, of which one contained mainly East Asian cultivars (L-EA), and the other (L-OT) showed no specificity for geographic distribution. Some ornamental cultivars (O-B) were included in clade L-EA. Note that two wild species (*P. davidiana* 2325-21A, and *P. mira* 2228-21A) were located near or in landrace clusters, probably because of recent frequent hybridization with domesticated peach cultivars or ancestral shared polymorphisms, as supported by PCA and STRUCTURE analyses (see below).

To examine the structure in more detail, PCA, population structure, and F_{ST} analysis were conducted for 87 *Amygdalus* subgenus

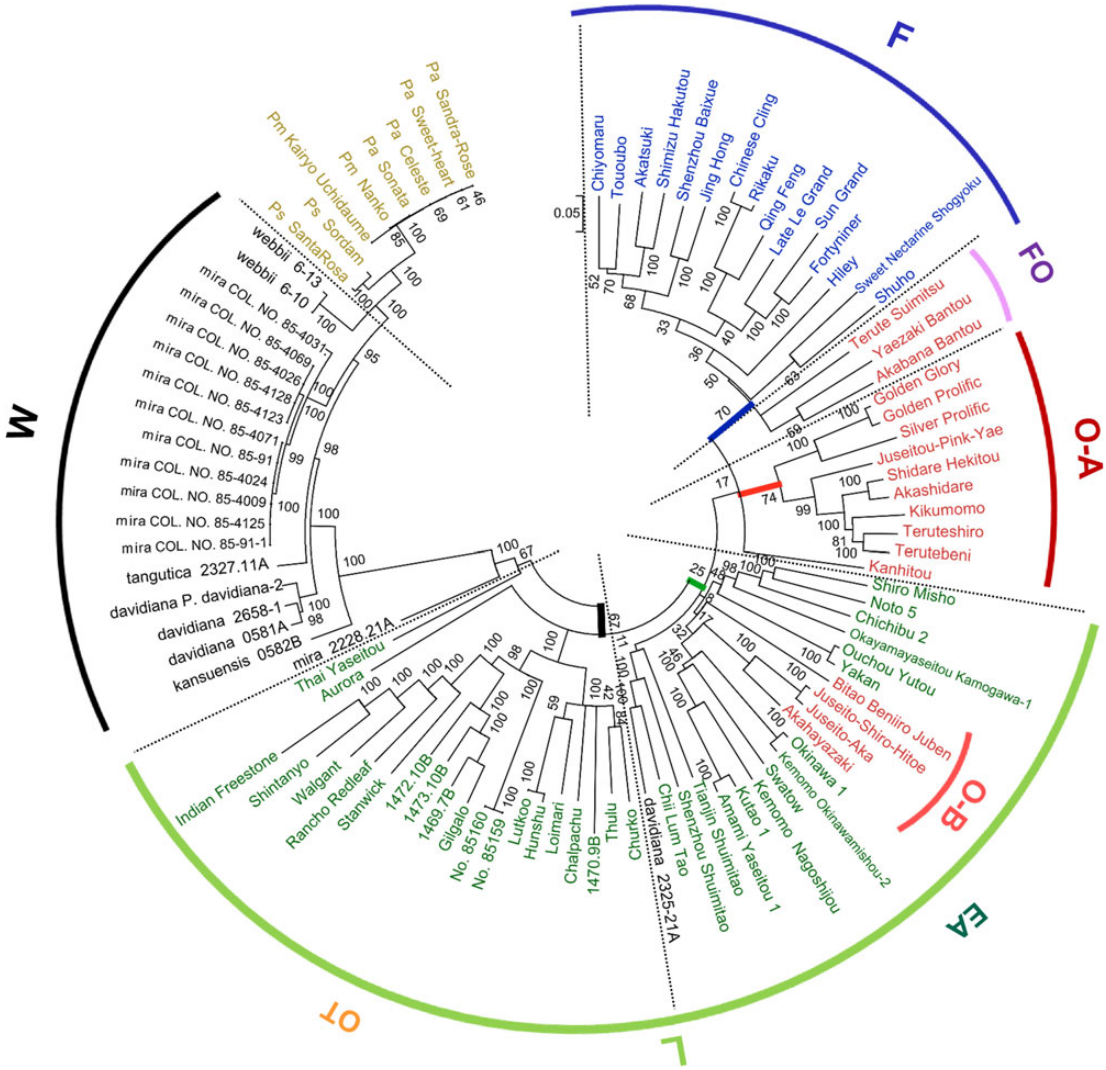

Figure 1. Evolutionary tree of domesticated peach and its wild relatives. Evolutionary tree constructed by a maximum likelihood (ML) approach using information from 5,180 genome-wide SNPs in 87 accessions of domesticated peach (*Prunus persica*) and five wild species (*P. davidiana*, *P. mira*, *P. kansuensis*, *P. tangutica*, and *P. webbii*), with 8 accessions of outgroup species (*P. salicina*, *P. mume*, and *P. avium*). Operational taxonomic units (OTU) are colored in ocher, black, green, blue, and red, for major clusters: wild species (W), landraces (L) (EA and OT for accessions in East Asia and other region, respectively), modern fruit cultivars (F), and modern ornamental cultivars (O-A and O-B for only ornamental usage, and FO for fruit and ornamentals), respectively, according to the tree topology, and to classical classifications considering uses and morphologies. The main differentiation steps of the domesticated peach, the East Asian cultivars (including modern cultivars), the O-A cluster, and the F cluster, are shown as black, green, red, and blue thick branches, respectively, with statistically significant support (bootstrap > 650/1,000, except for the differentiation to the East Asian cultivars).

accessions using the 5,180 SNP set. PCA analysis showed distinct differentiation of the W clusters from *P. persica* with two exceptions (*P. davidiana* 2325-21A and *P. mira* 2228-21A) (Fig. 2A for the first two principal components). In addition, this analysis revealed that the F, FO, and O-A clusters had genetic structures that were distinct from those of the others, although O-A may have experienced some hybridization with the landraces in East Asia or ancestrally shared polymorphisms with them (Fig. 2A). The results of population structure analysis by STRUCTURE 2.2 and InStruct, assuming subpopulation $K = 2$–10 distinct clusters, were consistent with each other, and supported almost the same conclusion as in the evolutionary tree and the PCA (Fig. 2B and C, for $K = 2$–7). Model selection based on the ΔK and ln Pr(K) supported $K = 6$ (Supplementary Fig. S1A and B). With $K = 5$ or less, ornamental cultivars and landraces were not clearly separated, whereas with $K = 7$, they displayed different population structures, although some genetic hybridization was likely (Fig. 2B and C, Supplementary Fig. S1C for comparison of the O and L-EA clusters in predefining two ancestral gene pools). The FO cluster was inferred to have experienced clear hybridization with the F and O/L-EA clusters in $K = 4$–7, a reasonable inference, given that they originated by recent hybridization between fruit and ornamental cultivars (Supplementary Fig. S1D). Two accessions of wild species, *P. davidiana* 2325-21A and *P. mira* 2228-21A, appeared to have experienced recent hybridization or ancestrally shared polymorphic alleles with landraces at $K = 7$, and for $K = 2$, their structures were almost the same as that of the domesticated *P. persica* (Fig. 2B and C). The F_{ST} test supported frequent hybridization in the FO cluster and the F and L-EA clusters and suggested significant differentiation in the domesticated peach (*P. persica*) from wild relatives ($F_{ST} = 0.268$ and 0.275 from *P. mira* and *P. davidiana*, respectively), and in F and O-A ($F_{ST} = 0.158$ and 0.135, respectively) from the landrace L-EA (Table 1), relative to the F_{ST} values for other tree crops and wild relatives.[18,22] Also, the O-A and O-B clusters, both of which are categorized as ornamental varieties, showed distinct differentiation ($F_{ST} = 0.228$).

Based on these population structure results, the O-B cluster was included in the L-EA (or L) cluster and the F-O was no longer considered because at least one of them originated from the recent hybridization between F and O-A gene pools (Supplementary Fig. S1D).

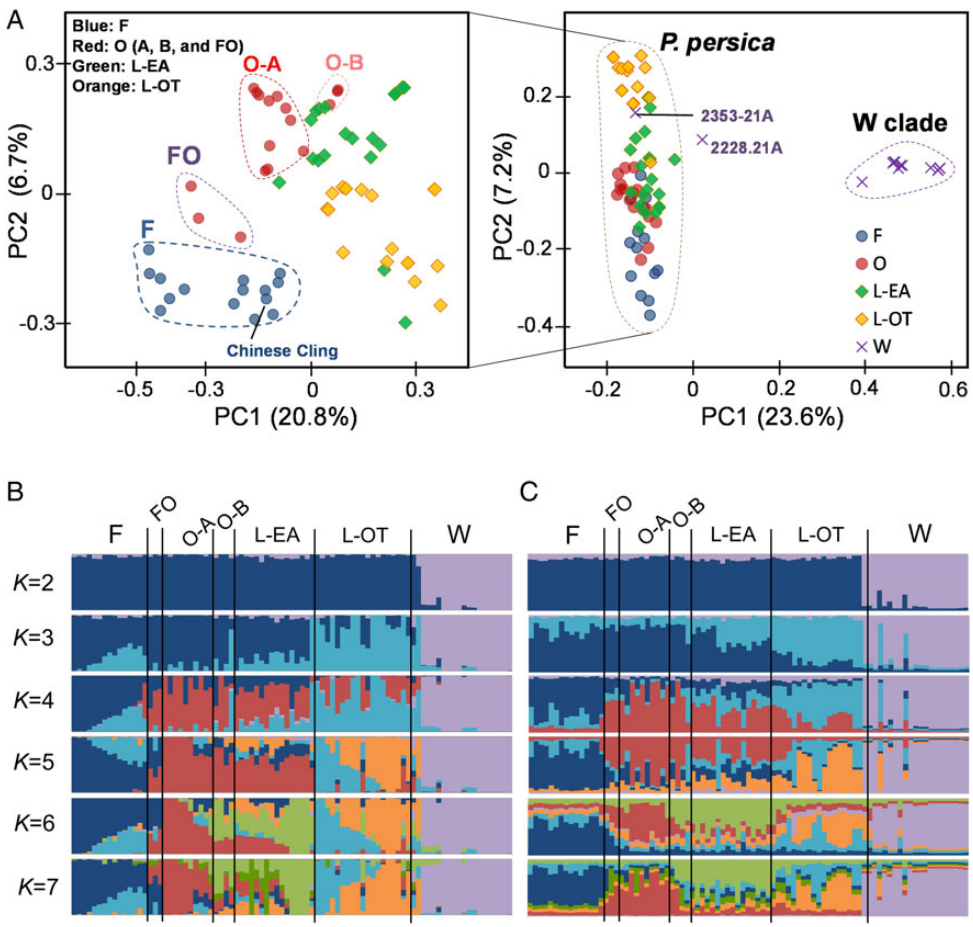

Figure 2. Population structure analysis in peach. (A) Principal component analysis using information from 5,180 genome-wide SNPs in 87 accessions including domesticated peach and its wild relatives (right) and in 67 accessions focusing on domesticated peach (left). The first two components in PCA (PC1 and PC2) are plotted on the axes to visualize the genetic relationships. The proportion of variance explained by each PC is given in parentheses along each axis. The wild species is shown as a cross. The landraces are shown as squares in green and yellow for accessions in East Asia and other regions, respectively. The modern cultivars are shown as circles in blue and red, corresponding to fruit and ornamental cultivars, respectively. (B and C) Structure analysis of subdivision of the population ($K = 2$–7), with STRUCTURE 2.2 (B) and with InStruct (C). Each individual is shown as a vertical bar. In $K = 2$, wild species showed a cluster distinct from domesticated peach, except for two accessions experiencing frequent hybridization with domesticated peach. In $K = 5$ or more, the F, L-OT, and O-A/O-B/L-EA clusters show clear separation. The O-A and O-B/L-EA showed significant separation at $K = 7$ or more.

Table 1. Pairwise F_{ST} in comparison to domesticated/wild species (A) and subclusters (B)

A	Peach P-AL	Wild relatives				
		W-AL	DAV	MIR		
Domesticated peach (*Prunus persica*) (P-AL)						
All wild relatives (W-AL)	0.301					
P. davidiana (DAV)	0.275	–				
P. mira (MIR)	0.268	–	0.212			
B	F	FO	O-A	O-B	L-EA	L-OT
Fruit cluster (F)						
Ornamental						
Fruit and ornamental cluster (O-F)	0.062					
Ornamental-A cluster (O-A)	0.245	0.141				
Ornamental-B cluster (O-B)	0.229	0.222	0.228			
Landraces						
East Asia (L-EA)	0.158	0.058	0.135	0.103		
Other regions (L-OT)	0.189	0.154	0.205	0.211	0.092	

3.2. Detection of genetic diversity, LD, and inbreeding/bottlenecks

Using the information from the 5,180 SNP set, we evaluated the possibility of inbreeding/bottleneck situations by comparing observed H_O versus H_E and using F_{IS} values with a null hypothesis of random mating within each species. When considering the entire domesticated peach group (*P. persica*, $n = 67$), the H_O value was found to be significantly lower than the He value ($H_O = 0.264$, $H_E = 0.401$, Supplementary Table S2), which suggests the occurrence of a population bottleneck or a recent shift to higher inbreeding during domestication. A similar situation was observed for the L clusters ($H_O = 0.220$, $H_E = 0.375$ for L-EA, $H_O = 0.173$, $H_E = 0.345$ for L-OT). These findings were also supported by F_{IS} values, suggesting degrees of inbreeding ($F_{IS} = 0.33$–0.52; Supplementary Table S2). In contrast, the F and O clusters demonstrated no such trends in H_O/H_E and F_{IS} values (Supplementary Table S2). Pairwise characterization of IBD to infer pedigree relationships among 67 *P. persica* individuals, however, showed that all accessions in the F cluster had first- or second-degree relationships with cv. Chinese Cling (pi-hat > 0.25 for all individuals, average pi-hat = 0.44; Supplementary Fig. S2), supporting previous proposals regarding the establishment of modern cultivars[32,33] and supporting the presence of strong genetic drift in the establishment of the F cluster. We detected significant IBD proportions between *P. davidiana* 2325-21A or *P. mira* 2228-21A and some individuals in the L cluster (pi-hat = ~0.189), despite the fact that *P. mira* and *P. davidiana* nested to the W cluster revealed no significant IBD values (pi-hat = 0). Moreover, the FO cluster demonstrated significant IBD values against the F and O-A clusters (pi-hat = ~0.419), although the F and O-A clusters have almost no pedigree relationships. These results support the possibility of recent hybridizations among these clusters.

The degree of population structuring is related to LD extension/decay. In domesticated peach, LD in European and US cultivars has been reported to be longer (up to 13–15 cM) than in other major perennial crops such as grape (<10 kb),[18] based on an analysis of 50 microsatellite markers by Aranzana et al.[33] However, the LD in landraces and the edible peaches primarily from China calculated using genome-wide SNPs was much shorter (up to 50 kb).[30] The LD decay in domesticated peach cultivars, in which the average pairwise LD values showed no significant difference against the background, was extended to ~1,000–2,500 kb and would be even more delayed in the wild relatives (although the LD decay in the Wild cluster was unclear, presumably owing to a shortage of effective accession numbers and of polymorphisms in the W cluster when this SNPs chip was used) (Fig. 3A for chromosome 4, Supplementary Fig. S3 for other chromosomes and LD plots). This distance of LD decay in the domesticated peach population is consistent with that in US and European peach cultivars[33] and greater than that in other major domesticated crops such as rice (up to 50–150 kb),[9,60] sorghum (up to ca. 150 kb),[42] and maize (up to 2 kb).[61] In comparing clusters, the F cluster showed much greater delay in decay of LD than the two subclusters of the L cluster (L-EA and L-OT), ranging over 3,000 kb on some chromosomes (Fig. 3B, Supplementary Fig. S3). This finding supports a bottleneck or shift to higher inbreeding in the establishment of the F cluster from the progenitor landraces. The results of IBD test would support a recent bottleneck from cv. Chinese Cling (Supplementary Fig. S2). The O-A cluster, in general, showed no significant differences in LD decay from the two L subclusters (Fig. 3B), although a slightly significant increase in LD could be detected for the O-A cluster on a few chromosomes (Supplementary Fig. S3). A significant delay in LD decay might also be expected from the distribution of EHH presented later. In the genome-wide LD analysis, we detected a clear monotonic decrease in static values of LD with the physical distance on any chromosome ranging from at least 1,000 kb in the domesticated peach cluster to even more in the F subcluster. This situation would make it preferable to identify selective sweeps based on LD index using small numbers of marker sets. In this study, we could use at least 5,000 pruned SNP sets (one SNP locus per *ca.* 45 kb on average in the peach genome) for subsequent analysis of selected genomic regions, yielding sufficient information for the identification of statistically distorted LD values from each SNP core in the genome.

3.3. Genome-wide detection of selection in domestication and cultivar differentiation pathways

To evaluate selective sweeps in local genomic regions, indexes based on the site frequency spectrum (SFS), such as π or Tajima's D which can detect fixed sweeps in a population, have been applied in plant species.[8,12–14,30] In genome-wide analysis of selected regions, similar approaches including LD- or haploblock-based methods, which usually detect very recent and ongoing sweeps, are reported to be effective for some plant species[9,10,62] and for some domesticated animals such as dog[63]

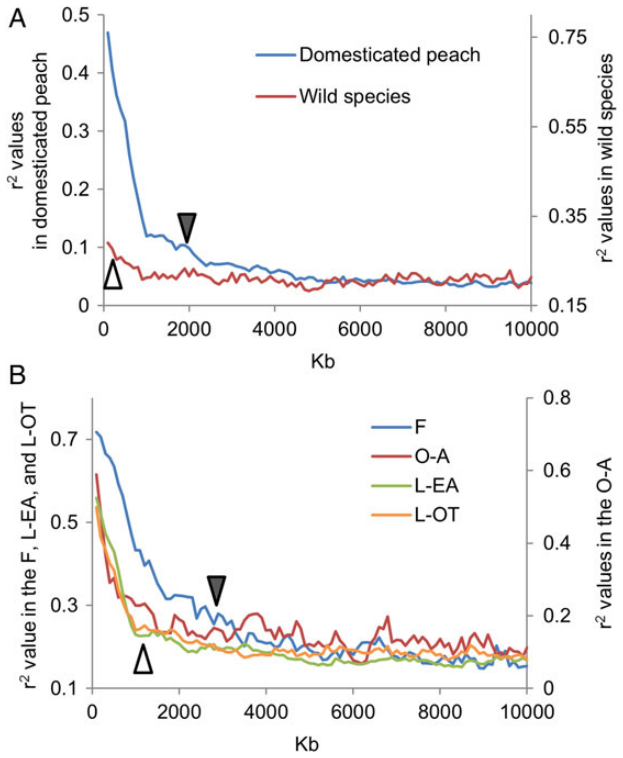

Figure 3. Comparison of LD decay in the wild species and each variety complex of domesticated peach. The average values of R^2 in pairwise LD among the 1,346 SNPs on Chromosome 4, which carries the most SNPs among the eight chromosomes, in windows of up to 10,000 kb are shown. The *Y*-axis standards are adjusted according to the background values in each subpopulation for visualization of LD decay for comparison among subpopulations. The *X*-axis shows physical distance among the SNPs, and the average R^2 values at 100-kb intervals are plotted. (A) Comparison of LD decay in domesticated peach and wild relatives. Black and white arrows indicate LD decay points in domesticated peach and wild relative, respectively. (B) Comparison of LD decay in the F, O-A, L-EA, and L-OT clusters. The F cluster shows only significantly expanded LD, as shown by the black arrow, in comparison to the other clusters, whose LD decays are shown with white arrows and black bars.

and chicken.[39] Here, selections for domesticated peaches were expected to be recent or ongoing, particularly for the F cluster. Peach is a species with a long generation time (at least 2–3 yrs to flowering) and one for which individual genotypes can be maintained by vegetative propagation for long time periods, suggesting that selected genes may have been maintained in heterozygous states. Thus, we adopted indexes based not only on SFS but also on LD or haploblock for exploiting selective sweep, following genome-wide analyses of selection in human genomes.[7,52–54]

In SFS-based analysis, we evaluated the transition of pooled heterozygosity (H_p) of each cluster using informative SNPs in a 400-kb sliding window with 100-kb steps, following a previous report on SFS-based analysis in chicken.[39] The distributions of observed H_p values and *Z*-score values of H_p (ZH_p), which considered coalescent effects by using the entire SNP sets in the peach genome, are shown in Fig. 4A for the domesticated peach population and two subpopulations, F and O-A. All showed some genomic regions containing candidates for selective sweeps with significant homozygous states in comparison to the whole genome ($P < 0.001$). However, the possibility that they are derived from differences in the substitution ratio in the peach genome, distortion of the availability in SNP patterns, or simple drift cannot be ruled out. Comparison of the H_p values between the F- or O-A clusters and the L cluster (Fig. 4B) revealed a significant bias in homozygous states among the clusters in specific genomic regions and suggested that at least some represent true selective sweeps that specifically occurred in the path to modern fruit or ornamental cultivar development. The middle of Chromosome 7 (*ca.* 8,000–8,800 kb) and the bottom of Chromosome 4 (*ca.* 29,200–29,600 kb) showed particularly clear tendencies to selective sweeps specific to the F and O-A subpopulations, respectively (Fig. 4C and D).

LD- or haploblock-based analyses were performed in accordance with the concept of EHH.[52] The program Sweep[53] was used to evaluate the iHS,[54] which quantifies the difference of EHH values around the selected locus in one population, and the Cross-Population EHH (XP-EHH),[53] which calculates EHH values from the same SNPs core between two populations using phased SNP strings. In both analyses, EHH values collected in 1,000-kb sections from each SNP locus showed significant reductions (*ca.* 0.05).

The iHS in domesticated peach, as well as the two subpopulations F and O-A, showed specific patterns of significant peaks ($P < 0.0001$) corresponding to putative selective sweeps (Fig. 5A, Supplementary Figs S5 and S6A). The patterns of the peaks for the selective sweep in iHS were considerably different from those in the SFS-based analysis (Fig. 4), perhaps because the iHS detects mainly selected alleles in the heterozygous state, whereas the SFS-based analysis is applicable mainly to the detection of homozygosity. Still, the same positions for putative selection in SFS-based and iHS analyses were selected on Chromosome 7 (*ca.* 8,000–8,800 kb) in the F cluster ($P < 0.00001$ in iHS). The XP-EHH analyses were performed in comparisons between domesticated peach and the W clusters, and between the F or O-A and the L clusters (Fig. 5B and Supplementary Fig. S6B) to analyse genes selected in the course of domestication and cultivar differentiation. In the XP-EHH analysis of domesticated peach and the W clusters, the peaks were at the bottoms of Chromosome 4 (*ca.* 25,000–27,000 kb for $P < 0.001$), and Chromosome 6 (*ca.* 25,000–26,500 kb for $P < 0.001$). In XP-EHH analyses, they were similar to those in iHS analyses of the domesticated peach (Fig. 5, $P < 0.000001$ for chr. 4, and $P < 0.0001$ for chr. 6). Some of the significant peaks in XP-EHH analyses were different from those in the iHS and SFS-based approaches. Similar situations have been reported in analyses using EHH- and SFS-based methods in human evolution, which captured certain selected regions in all approaches, further supporting the value of using multiple approaches for comprehensive analysis of positive selection.[52,53] The XP-EHH analyses of the F or O-A cluster relative to the L cluster also showed some specific peaks, though different from the iHS and SFS-based results. Significant peaks could still be detected on Chromosome 4 (*ca.* 8,500–9,500 kb) and Chromosome 7 (*ca.* 8,000–8,800 kb) in the F cluster ($P < 0.00001$). The peak on Chromosome 4 was also detected in the domesticated peach cluster by the SFS-based method (Fig. 4).

3.4. Candidate genes under selection

For the F and O-A clusters, genetic diversity of candidate genes located on representative regions under putative strong positive selection was estimated. These included Chromosome 7 (8,000–8,800 kb) for the F cluster, Chromosomes 4 (29,200–29,400 kb) and 8 (16,900–17,550 kb) for the O-A cluster, and Chromosome 1 for both (1,200–1,500 kb), using the 30 domesticated peach accessions (Supplementary Table S3). Note that the focus was only on genes annotated with functions possibly conferring advantages of differentiation. Of these, ppa004528m, ppa016246m, and ppa025156m showed significant reduction in genetic diversity in the F cluster ($\pi = 0$–0.00007) against the O-A and

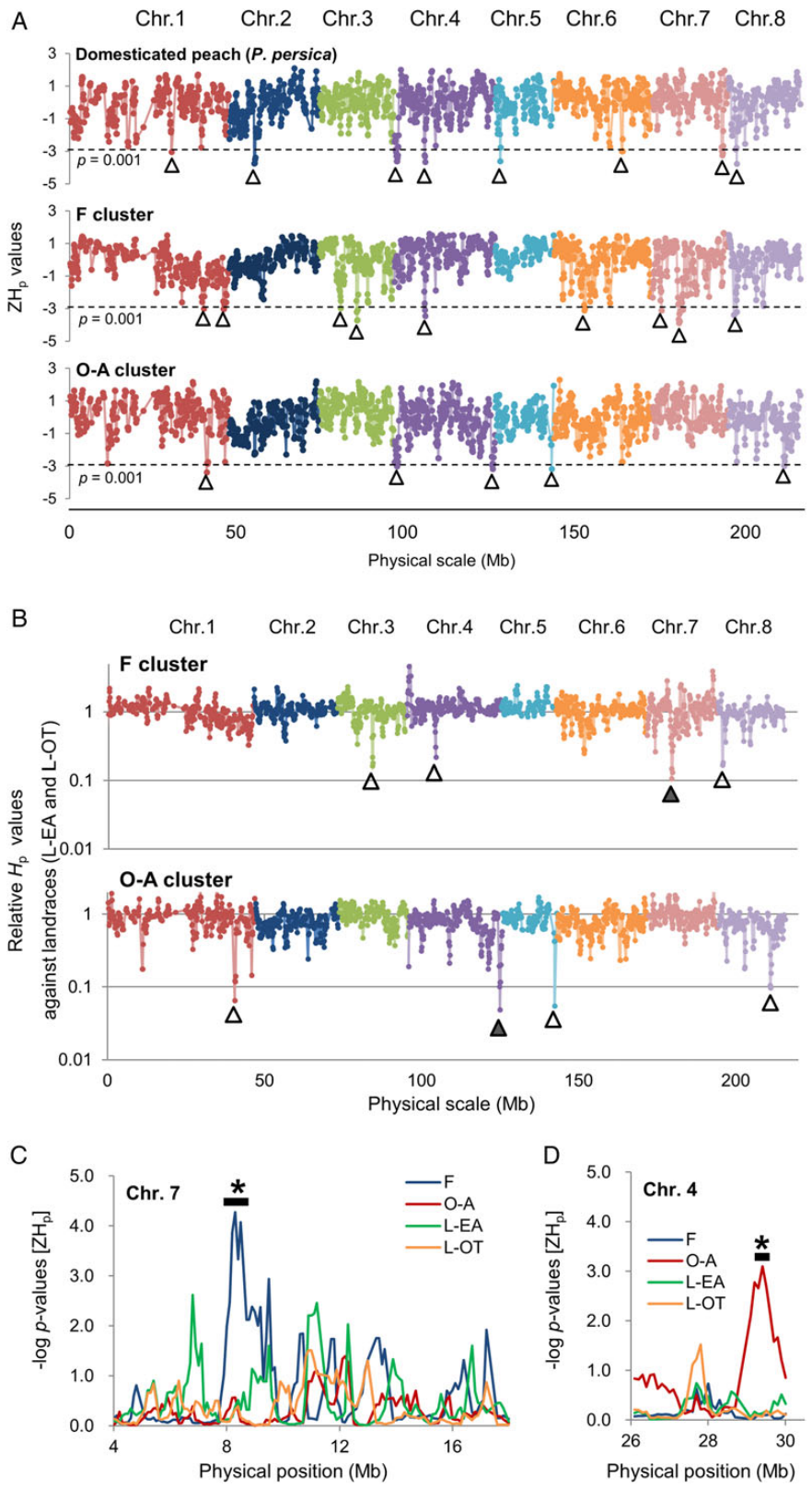

Figure 4. Genome-wide selective sweep analysis based on ZH_p, in the domesticated peach and two modern varietal complexes. (A) Transitions of Z-transformed values of H_p in 400-kb bins with 100-kb steps are shown for the domesticated peach population and two subpopulations, the F and O-A clusters. Putative regions showing selective sweeps ($P < 0.001$) are indicated by outlined triangles. (B) For the F and O-A clusters, plots of relative values of H_p in 400-kb bins with 100-kb steps against the L cluster are shown. Putative regions showing selective sweeps in the paths of differentiation from landraces are indicated by triangles. For the two regions indicated by black triangles, detailed characterization of ZH_p in a comparison among four subpopulations (F, O-A, L-EA, and L-OT) is shown in (C) for the top of Chromosome 7, which corresponded to a putative selective sweep in the F cluster, and in (D) for the bottom of Chromosome 4, showing a putative selective sweep in the O-A cluster.

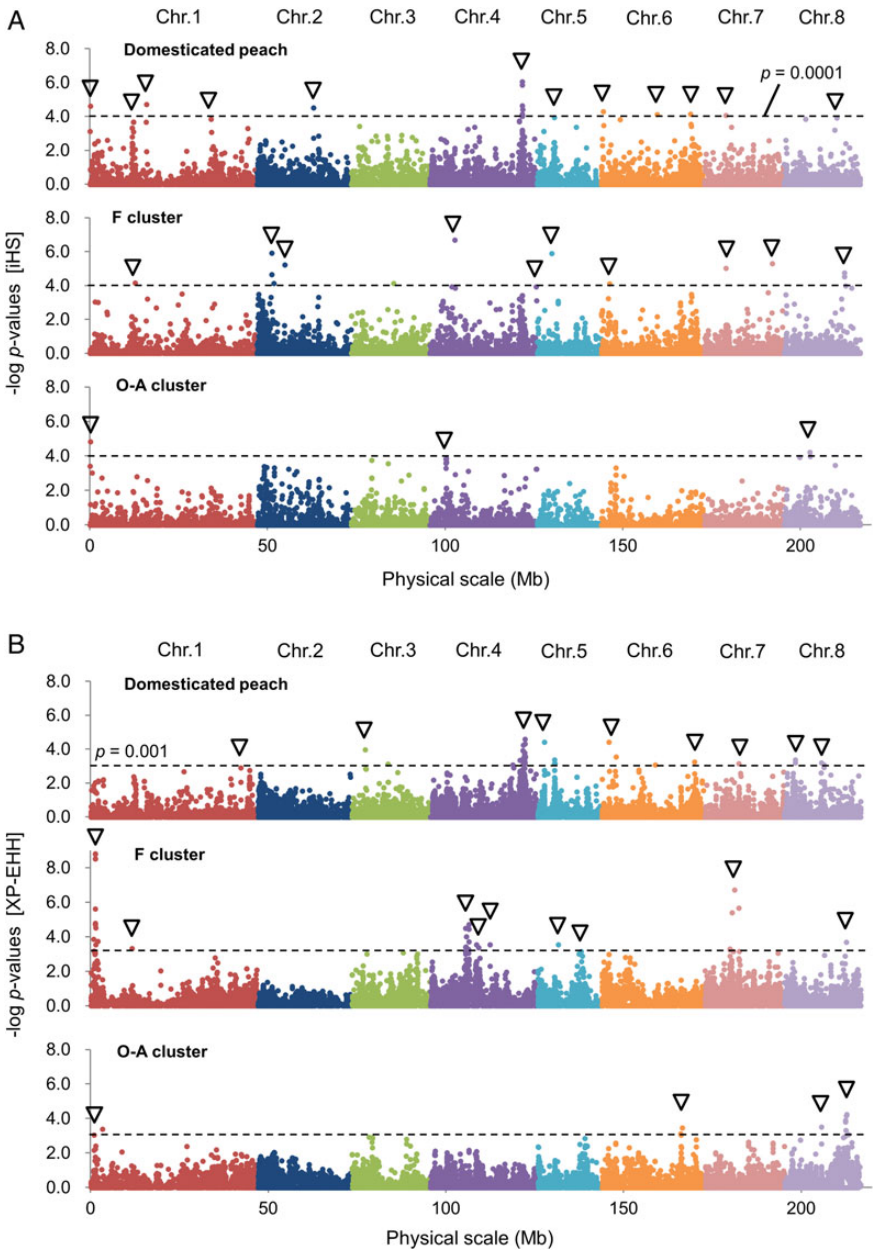

Figure 5. Genome-wide analysis for positive selection based on EHH. (A) iHS from each SNP core showing heterozygous states in one population transformed to *P* values with *R* and plotted as a logarithmic value. Underlining indicates threshold values for selection ($P < 0.0001$). (B) XP-EHH from each SNP core showing the same nucleotide between the subject and the comparison target, also transformed to *P* values and plotted in logarithmic scale. For the domesticated peach, the wild relative cluster was used for comparison. For the F and O-A clusters, the landrace (L) cluster was independently used for comparison. Underlining indicates threshold values for selection ($P < 0.001$). In both iHS and XP-EHH tests, putative selected regions are indicated by outlined triangles.

landrace (L-EA and L-OT) clusters (average $\pi = 0.00114 \pm 0.00021$, 0.00279 ± 0.00011, and 0.00629 ± 0.00091, respectively, Supplementary Table S4). However, no disruptions or substitutions clearly associated with the functional changes in the ORF sequences could be identified, although minor substitutions on protein sequences were observed in comparison to some landraces. For the O-A cluster, ppa003808m showed a significant reduction in genetic diversity ($\pi = 0$) in comparison to the other clusters (average $\pi = 0.00150 \pm 0.00017$, Supplementary Table S4). ppa003808m retained a trihelix motif, showing high similarity to the *PETALLOSS* (*PTL*) gene, which is involved in petal size and morphology in *Arabidopsis thaliana*.[64–66]

4. Discussion

4.1. Domestication path in peach

The domestication events for major annual crops, such as maize or rice, have been well characterized, owing in part to the relatively straightforward selection for consumable parts as well as the genetic fixation of key controlling genes in distinct species or cultivars. Tree crops appear to have more complicated domestication histories.[18,22] For example, landrace accessions of *P. persica*, such as Kemomo Nagoshijou or Kutao, which show wild species-like characteristics, share a similar genetic structure with cultivated peach accessions. Thus, it

may be difficult to define a single domestication path. Two scenarios are supported by the data: (i) current wild peach cultivars (feral peaches) were derived from divergent lines after a general domestication event or (ii) multiple independent domestications occurred for each trait or region. The first scenario (i) would fit well with the premise that *P. persica* was first domesticated from a wild species in China before diffusion to other regions.[28,29] Such a general domestication event is supported by the tendencies for population bottlenecks or inbreeding to be detected using the full population of peach (*P. persica*) in this study. In contrast, other domesticated perennial crops such as grape and apple are reported to lack narrow domestication bottlenecks.[18,22,67] However, in theory, perennial woody crops are under strong bottleneck pressures during domestication, resulting from the ease of their clonal propagation as well as difficulties associated with their long generation time. The wide domestication bottleneck in grape might be explained by widespread use of vegetative propagation of a large number of landraces with a resulting maintenance of genetic diversity as well as of many local genetic clusters.[18] This situation would be more aligned with the second scenario (ii). The results of IBD analysis in peach showed that, except for the F cluster, most cultivars have no first- or second-degree relationships. This situation is distinct from that of domesticated grapevine, which shows a high ratio of first-degree relationships with a relatively small number of representative cultivars.[18] This finding indicates that there were not many opportunities for domestication or construction of local clusters in peach, at least for the cultivars examined in this study.

4.2. Selection in peach domestication and cultivar differentiation

Ancient selection in the peach genome appears likely, although some observations may be due to genetic drift. Results from QTL analysis have also been used to support selection during domestication and cultivar differentiation in rice[8] and chicken[39] In the present study, the strong putative selection on Chromosome 4 (*ca* 8,500–9,500 kb) detected in modern fruit cultivars (the F cluster) is consistent with a major QTL (UDP96-003) for important domestication traits, including fruit size and maturity date previously reported by Quilot et al. in a backcross between cultivated peach and its wild relative *P. davidiana*.[68] This finding suggests that favourable allele(s) in this region had been selected from wild ancestors during domestication. Furthermore, evidence for selection is observed at the bottom of Chromosome 6 (*ca.* 25,000–26,500 kb) which corresponds to the genomic regions XP-EHH and iHS, known to contain the self-incompatibility (*S*) controlling haplotype, including an F-box gene as the pollen-*S* (*SFB*) factor and an RNase gene as the pistil-*S* (*S*-RNase) factor in *Prunus* species.[69,70] Previously, it has been suggested that the whole domesticated peach (*P. persica*), including all clusters apart from the W cluster in this study, is a self-compatible (SC) species with at least four forms of pollen-*S*, $PpSFB_{1–4}$, disruptive mutations.[71,72] Meanwhile, to some extent, the wild species (W cluster) in the *Amygdalus* subgenus are supposed to demonstrate SI.[73] In general, a shift from SI to SC is strongly selected during domestication to facilitate the stable production of inbred lines.[27] In contrast, a change in the mating system from SI to SC could have a strong influence on the pattern of polymorphisms, affecting genetic diversity and LD. This might result in the differences in genetic structures between wild species and domesticated peach.

At least three candidates for selection were identified in the F cluster, with many on Chromosome 7. The candidates, ppa004528m, ppa016246m, and ppa025156m, show high similarity to *lycopene cyclase*-like At3g10230 (*LYC*)/At5g57030 (*LUT2*), At3g18030 (*Arabidopsis thaliana Hal3-like protein A*; *AtHAL3A*), and At1g18990 (*reduced vernalization response 1*; *VRN1*) in the *Arabidopsis thaliana* genome. Lycopene cyclases (LYCs) play a role in the biosynthesis of lutein, which is a member of the carotenoid pathway.[74] The simplest explanation for the selection of ppa004528m may be its association with the yellow carotenoid pigmentation in peach. However, the flesh coloration of peach is determined mainly by the *Y* locus, which is reportedly located on Chromosome 1 (LG1).[75,76] Major QTLs for skin color are also different from those in most regions identified as having experienced selective sweeps, although one of them is located on Chromosome 7.[77] The other two candidates, *AtHal3A*-like and *VRN1*-like genes, may contribute salt/osmotic tolerance and flowering/bud-burst timing, respectively, based on the functions of their homologs in *Arabidopsis*.[78,79] In considering the uses and characteristics of modern fruit cultivars, we may expect that some genes directly involved in fruit traits, such as fruit size, sugar contents, or the important flesh-softening trait often called 'melting texture', have been under strong selection. Where selection for more ornamental traits is expected, as in the O-A cluster, ppa003808m appears to be a good candidate, given that the associated *PETALLOSS* trait is known to affect morphological or architectural changes, particularly for flowers.

Genes with annotation for fruit quality and ornamental value are limited at present, presumably because of the lack of information for these traits in model plant species. Thus, the still-uncharacterized genes located in the selected regions in the F and O-A clusters may provide opportunities for elucidating critical tree domestication factors not present in current annual plant models. Genes controlling such key domestication and cultivar differentiation factors could be identified by improved characterization of the selection pathways involved, perhaps by further exploiting the wide genetic diversity present in peach and similar perennial tree crops. This research approach could prove powerful, particularly for long-lived perennial crops where mutagenetic and map-based approaches are restricted.

Acknowledgements

We thank Dr Masanori Yamasaki, Food Resources Education and Research Center, Graduate School of Agricultural Science, Kobe University (Kasai, Hyogo 675-2103, Japan) for many critical suggestions on experimental design, and to Drs Luca Comai and Isabelle M. Henry, Genome Center, University of California Davis (Davis, CA, USA) for bioinformatic support. We gratefully acknowledge gifts of plant materials from Yukio Sasabe, Research Institute for Agriculture Okayama Prefectural Technology Center for Agriculture, Forestry, and Fisheries (Akaiwa, Japan).

Supplementary data

Supplementary data are available at www.dnaresearch.oxfordjournals.org.

Funding

This work was supported by a Grant-in-Aid for Research Activity Start-up from the Japan Society for the Promotion of Science (number 23880013 to T.A.). Funding to pay the Open Access publication charges for this article was provided by Grant-in-Aid for Young Scientists (A) (no. 26712005 to TA) from JSPS.

References

1. Zeder, M.A., Emshwiller, E., Smith, B.D. and Bradley, D.G. 2006, Documenting domestication: the intersection of genetics and archaeology, *Trends Genet.*, **22**, 139–55.

2. Doebley, J.F., Gaut, B.S. and Smith, B.D. 2006, The molecular genetics of crop domestication, *Cell*, **127**, 1309–21.
3. Caicedo, A.L., Williamson, S.H., Hernandez, R.D., et al. 2007, Genome-wide patterns of nucleotide polymorphism in domesticated rice, *PLoS Genet.*, **3**, e163.
4. Allaby, R.G., Fuller, D.Q. and Brown, T.A. 2008, The genetic expectations of a protracted model for the origins of domesticated crops, *Proc. Natl Acad. Sci. USA*, 107, **105**, 13982–86.
5. Gross, B.L. and Olsen, K.M. 2010, Genetic perspectives on crop domestication, *Trends Plant Sci.*, **15**, 529–37.
6. Yamasaki, M., Wright, S.I. and McMullen, M.D. 2007, Genomic screening for artificial selection during domestication and improvement in maize, *Ann. Bot.*, **100**, 967–73.
7. Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C. and Clark, A.G. 2007, Recent and ongoing selection in the human genome, *Nat. Rev. Genet.*, **8**, 857–68.
8. Asano, K., Yamasaki, M., Takuno, S., et al. 2011, Artificial selection for a green revolution gene during japonica rice domestication, *Proc. Natl Acad. Sci. USA*, **108**, 11034–39.
9. Xu, X., Liu, X., Ge, S., et al. 2012, Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes, *Nat. Biotech.*, **30**, 105–11.
10. Cavanagh, C.R., Chao, S., Wang, S., et al. 2013, Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars, *Proc. Natl Acad. Sci. USA*, **110**, 8057–62.
11. Clark, R.M., Linton, E., Messing, J. and Doebley, J.F. 2004, Pattern of diversity in the genomic region near the maize domestication gene *tb1*, *Proc. Natl Acad. Sci. USA*, **101**, 700–7.
12. Palaisa, K., Morgante, M., Tingey, S. and Rafalski, A. 2004, Long-range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep, *Proc. Natl Acad. Sci. USA*, **101**, 9885–90.
13. Wright, S.I., Bi, I.V., Schroeder, S.G., et al. 2005, The effects of artificial selection on the maize genome, *Science*, **308**, 1310–4.
14. Yamasaki, M., Tenaillon, M.I., Bi, I.V., et al. 2005, A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement, *Plant Cell*, **17**, 2859–72.
15. Morrell, P.L., Buckler, E.S. and Ross-Ibarra, J. 2011, Crop genomics: advances and applications, *Nat. Rev. Genet.*, **13**, 85–96.
16. Heffner, E.L., Sorrells, M.E. and Jannink, J.L. 2009, Genomic selection for crop improvement, *Crop Sci.*, **49**, 1–12.
17. McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B. and Little, J. 2008, Genome-wide association studies for complex traits: consensus, uncertainty and challenges, *Nat. Rev. Genet.*, **9**, 356–9.
18. Myles, S., Boyko, A.R., Owens, C.L., et al. 2011, Genetic structure and domestication history of the grape, *Proc. Natl Acad. Sci. USA*, **108**, 3530–5.
19. Chen, H., Morrell, P.L., Ashworth, V.E.T.M., de la Cruz, M. and Clegg, M. T. 2009, Tracing the geographic origins of major avocado cultivars, *J. Hered.*, **100**, 56.65.
20. Miller, A. and Schaal, B. 2005, Domestication of a Mesoamerican cultivated fruit tree, *Spondias purpurea*, *Proc. Natl Acad. Sci. USA*, **102**, 12801–06.
21. Miller, A. and Gross, B.L. 2011, From forest to field: perennial fruit crops domestication, *Am. J. Bot.*, **98**, 1389–414.
22. Cornille, A., Gladieux, P., Smulders, M.J., et al. 2012, New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties, *PLoS Genet.*, **8**, e1002703.
23. Austerlitz, F., Mariette, S., Machon, N., Gouyon, P.H. and Godelle, B. 2000, Effects of colonization processes on genetic diversity: differences between annual plants and tree species, *Genetics*, **154**, 1309–21.
24. Petit, R.J. and Hampe, A. 2006, Some evolutionary consequences of being a tree, *Annu. Rev. Ecol. Evol. Syst.*, **37**, 187–214.
25. Savolainen, O. and Pyhäjärvi, T. 2007, Genomic diversity in forest trees, *Curr. Opin. Plant Biol.*, **10**, 162–7.
26. Glémin, S. and Bataillon, T. 2009, A comparative view of the evolution of grasses under domestication, *New Phytol.*, **183**, 273–90.
27. Rowlands, D.G. 1964, Self-incompatibility in sexually propagated cultivated plants, *Euphytica*, **13**, 157–62.
28. Faust, M. and Timon, B. 1995, In: Janick, J. (ed) *Origin and dissemination of peach, Hort. Rev.*, Vol. **17**, Wiley: New York, pp. 331–80.
29. Huang, H., Cheng, Z., Zhang, Z. and Wang, Y. 2008, In: Layne, D.R. and Bassi, D. (eds) *History of cultivation and trends in China, the peach: botany production and uses*, CAB International, pp. 37–60.
30. Cao, K., Zheng, Z., Wang, L., et al. 2014, Comparative population genomics reveals the domestication history of the peach, *Prunus persica*, and human influences on perennial fruit crops, *Genome Biol.*, **15**, 415.
31. Scorza, R., Mehlenbacher, S.A. and Lightner, G.W. 1985, Inbreeding and coancestry of freestone peach cultivars of the eastern United States and implications for peach germplasm improvement, *J. Amer. Soc. Hort. Sci.*, **110**, 547–52.
32. Warburton, M.L. and Bliss, F.A. 1996, Genetic diversity in peach (*Prunus persica* L. Batch) revealed by randomly amplified polymorphic DNA (RAPD) markers and compared to inbreeding coefficients, *J. Am. Soc. Hortic. Sci.*, **121**, 1012–9.
33. Aranzana, M.J., Abbassi, E., Howad, W. and Arús, P. 2010, Genetic variation, population structure and linkage disequilibrium in peach commercial varieties, *BMC Genet.*, **11**, 69.
34. Cheng, Z.P. 2007, Genetic characterization of different demes in *Prunus persica* revealed by RAPD markers, *Sci. Hortic.*, **111**, 242–7.
35. Cheng, Z.P. and Huang, H. 2009, SSR fingerprinting Chinese peach cultivars and landraces (*Prunus persica*) and analysis of their genetic relationships, *Sci. Hortic.*, **120**, 188–93.
36. Xie, R., Li, X., Chai, M., et al. 2010, Evaluation of the genetic diversity of Asian peach accessions using a selected set of SSR markers, *Sci. Hortic.*, **125**, 622–9.
37. Martinez-Gómez, P., Arulsekar, S., Potter, D. and Gradziel, T.M. 2003, Relationships among peach, almond, and related species as detected by simple sequence repeat marker, *J. Am. Soc. Hortic. Sci.*, **128**, 667–71.
38. Hu, D., Zhang, Z., Zhang, D., Zhang, Q. and Li, J. 2005, Genetic relationship of ornamental peach determined using AFLP markers, *Hort. Sci.*, **40**, 1782–6.
39. Rubin, C.J., Zody, M.C., Eriksson, J., et al. 2010, Whole-genome resequencing reveals loci under selection during chicken domestication, *Nature*, **464**, 587–91.
40. Toomajian, C., Hu, T.T., Aranzana, M.J., et al. 2006, A nonparametric test reveals selection for rapid flowering in the Arabidopsis genome, *PLoS Biol.*, **4**, e137.
41. Hufford, M.B., Xu, X., van Heerwaarden, J., et al. 2012, Comparative population genomics of maize domestication and improvement, *Nat. Genet.*, **44**, 808–11.
42. Morris, G.P., Ramu, P., Deshpande, S.P., et al. 2013, Population genomic and genome-wide association studies of agroclimatic traits in sorghum, *Proc. Natl Acad. Sci. USA*, **110**, 453–8.
43. Verde, I., Bassil, N., Scalabrin, S., et al. 2012, Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm, *PLoS ONE*, **7**, 10.
44. International Peach Genome Initiative, Verde, I., Abbott, A.G., et al. 2013, The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution, *Nat. Genet.*, **45**, 487–94.
45. Purcell, S., Neale, B., Todd-Brown, K., et al. 2007, PLINK: a tool set for whole-genome association and population-based linkage analyses, *Am. J. Hum. Genet.*, **81**, 559–75.
46. Raymond, M. and Rousset, F. 1995, GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism, *J. Hered.*, **86**, 248–9.
47. Rousset, F. 2008, Genepop'007: a complete reimplementation of the Genepop software for Windows and *Linux*, *Mol. Ecol. Resources*, **8**, 103–6.
48. Pritchard, J.K., Stephens, M. and Donnelly, P. 2000, Inference of population structure using multilocus genotype data, *Genetics*, **155**, 945–59.
49. Gao, H., Williamson, S. and Bustamante, C.D. 2007, An MCMC approach for joint inference of population structure and inbreeding rates from multilocus genotype data, *Genetics*, **176**, 1635–51.
50. Evanno, G., Regnaut, S. and Goudet, J. 2005, Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study, *Mol. Ecol.*, **14**, 2611–20.

51. Scheet, P. and Stephens, M. 2006, A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase, *Am. J. Hum. Genet.*, **78**, 629–44.
52. Sabeti, P.C., Reich, D.E., Higgins, J.M., et al. 2002, Detecting recent positive selection in the human genome from haplotype structure, *Nature*, **419**, 832–7.
53. Sabeti, P.C., Varilly, P., Fry, B., et al. 2007, Genome-wide detection and characterization of positive selection in human populations, *Nature*, **449**, 913–8.
54. Voight, B.F., Kudaravalli, S., Wen, X. and Pritchard, J.K. 2006, A map of recent positive selection in the human genome, *PLoS Biol.*, **4**, e72.
55. Akagi, T., Henry, I.M., Tao, R. and Comai, L. 2014, A Y-chromosome–encoded small RNA acts as a sex determinant in persimmons, *Science*, **346**, 646–50.
56. Li, H. and Durbin, R. 2009, Fast and accurate short read alignment with Burrows-Wheeler transform, *Bioinformatics*, **25**, 1754–60.
57. Li, H., Handsaker, B., Wysoker, A., et al. 2009, The sequence alignment/map format and SAMtools, *Bioinformatics*, **25**, 2078–9.
58. Librado, P. and Rozas, J. 2009, DnaSP v5: a software for comprehensive analysis of DNA polymorphism data, *Bioinformatics*, **25**, 1451–2.
59. Gouy, M., Guindon, S. and Gascuel, O. 2010, SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building, *Mol. Biol. Evol.*, **27**, 221–4.
60. Mather, K.A., Caicedo, A.L., Polato, N.R., et al. 2007, The extent of linkage disequilibrium in rice (*Oryza sativa* L.), *Genetics*, **177**, 2223–32.
61. Yan, J., Shah, T., Warburton, M.L., et al. 2009, Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers, *PLoS ONE*, **4**, e8451.
62. Hancock, A.M., Brachi, B., Faure, N., et al. 2011, Adaptation to climate across the *Arabidopsis thaliana* genome, *Science*, **334**, 83–6.
63. Akey, J.M., Ruhe, A.L., Akey, D.T., et al. 2010, Tracking footprints of artificial selection in the dog genome, *Proc. Natl Acad. Sci. USA*, **107**, 1160–5.
64. Griffith, M.E., da Silva Conceição, A. and Smyth, D.R. 1999, *PETALLOSS* gene regulates initiation and orientation of second whorl organs in the *Arabidopsis* flower, *Development*, **126**, 5635–44.
65. Brewer, P.B., Howles, P.A., Dorian, K., et al. 2004, *PETALLOSS*, a trihelix transcription factor gene, regulates perianth architecture in the *Arabidopsis* flower, *Development*, **131**, 4035–45.
66. Lampugnani, E.R., Kilinc, A. and Smyth, D.R. 2012, *PETALLOSS* is a boundary gene that inhibits growth between developing sepals in *Arabidopsis thaliana*, *Plant J.*, **71**, 724–35.
67. Grassi, F., Labra, M., Imazio, S., et al. 2003, Evidence of a secondary grapevine domestication centre detected by SSR analysis, *Theor. Appl. Genet.*, **107**, 1315–20.
68. Quilot, B., Wu, B.H., Kervella, J., et al. 2004, QTL analysis of quality traits in an advanced backcross between *Prunus persica* cultivars and the wild relative species, *P. davidiana*. *Theor. Appl. Genet.*, **109**, 884–97.
69. Ushijima, K., Sassa, H., Dandekar, A.M., et al. 2003, Structural and transcriptional analysis of the self-incompatibility locus of almond: identification of a pollen-expressed F-box gene with haplotype-specific polymorphism, *Plant Cell*, **15**, 771–81.
70. Tao, R. and Iezzoni, A.F. 2010, The *S*-RNase-based gametophytic self-incompatibility system in *Prunus* exhibits distinct genetic and molecular features, *Sci. Hort.*, **124**, 423–33.
71. Tao, R., Watari, A., Hanada, T., et al. 2007, Self-compatible peach (*Prunus persica*) has mutant versions of the *S* haplotypes found in self-incompatible *Prunus* species, *Plant Mol. Biol.*, **63**, 109–23.
72. Hanada, T., Watari, A., Kibe, T., et al. 2014, Two novel self-compatible *S* haplotypes in peach (*Prunus persica*), *J. Jpn Soc. Hortic. Sci.*, **83**, 201–13.
73. Banović, B., Šurbanovski, N., Konstantinović, M. and Maksimović, V. 2009, Basic RNases of wild almond (*Prunus webbii*): cloning and characterization of six new *S*-RNase and one "non-*S* RNase" genes, *J. Plant Physiol.*, **166**, 395–402.
74. Giuliano, G., Tavazza, R., Diretto, G., Beyer, P. and Taylor, M.A. 2008, Metabolic engineering of carotenoid biosynthesis in plants, *Trends Biotech.*, **26**, 139–45.
75. Bliss, F.A., Arulsekar, S., Foolad, M.R., et al. 2002, An expanded genetic linkage map of *Prunus* based on an interspecific cross between almond and peach, *Genome*, **45**, 520–9.
76. Warburton, M.L., Becerra-Velásquez, V.L., Goffreda, J.C. and Bliss, F.A. 1996, Utility of RAPD markers in identifying genetic linkages to genes of economic interest in peach, *Theor. Appl. Genet.*, **93**, 920–5.
77. Eduardo, I., Pacheco, I., Chietera, G., et al. 2011, QTL analysis of fruit quality traits in two peach intraspecific populations and importance of maturity date pleiotropic effect, *Tree Genet. Genomes*, **7**, 323–35.
78. Espinosa-Ruiz, A., Bellés, J.M., Serrano, R. and Culiáñez-Macià, F.A. 1999, *Arabidopsis thaliana* AtHAL3: a flavoprotein related to salt and osmotic tolerance and plant growth, *Plant J.*, **20**, 529–39.
79. Levy, Y.Y., Mesnage, S., Mylne, J.S., Gendall, A.R. and Dean, C. 2002, Multiple roles of *Arabidopsis VRN1* in vernalization and flowering time control, *Science*, **297**, 243–6.